Research Journal in Advanced Humanities

RJAH

https://doi.org/10.58256/5epj5m33

Open Access

Check for updates

REVIEW ARTICLE

Section: *Literature, Linguistics & Criticism*

Discourse analysis of cyberbullying among university students: An applied linguistics approach

Norah Nasser Saleh Alowayyid[1]*, Mohamad Ahmad Saleem Khasawneh[2]

[1]Department of Educational Sciences, Prince Sattam bin Abdulaziz University, Saudi Arabia

[2]Special Education Department, King Khalid University, Saudi Arabia

*Correspondence: n.alowayyid@psau.edu.sa

ABSTRACT

Cyberbullying in university contexts is often investigated as a behavioural, psychological, or institutional safety phenomenon. A few studies have comparatively conceptualized it as a discursive practice as power, identity, and group belonging are examined in complicated interactions. Adopting applied-linguistics as an approach, this paper advances an integrated framework for analyzing cyberbullying as discursive harm—a patterned configuration of stance acts, evaluative positioning, and face-attacking moves that scale across turns, participants, and semiotic modes. Theoretically, we synthesize Computer-Mediated Discourse Analysis (Herring, 2004), Critical Discourse Analysis (CDA) grounded in Fairclough's three-dimensional logic (as operationalized by Janks, 2005), Appraisal theory (Martin & White, 2005), and facework/(im)politeness (Goffman, 1967; Brown & Levinson, 1987; Culpeper, 1996) to propose a novel model: The Stance–Face–Power (SFP) Model. Methodologically, we introduce Corpus-Assisted, Interactionally-Refined CDA (CAIR-CDA), which couples corpus-assisted discourse studies procedures (pattern discovery at scale) with sequential, turn-by-turn interactional analysis and multimodal annotation, enabling analysts to connect micro-linguistic and semiotic choices (e.g., pronouns, modality, negation, mock politeness, evaluative lexis, evidential framing, emoji/meme uptake) to macro-social dynamics (e.g., gendered norm enforcement, institutional authority, group dominance, and platform affordances). A distinctive methodological contribution is the Dialogic Harm Index (DHI)—a transparent, reproducible metric integrating evaluative polarity/intensity (Appraisal), impoliteness strategy type, alignment and piling-on dynamics, and interactional power cues (commands, gatekeeping, exposure threats). To demonstrate analytic payoffs without reproducing harmful content or revealing identities, we provide a set of ethically "sanitized" composite vignettes and show how CAIR-CDA distinguishes cyberbullying from adjacent practices (e.g., conflict, critique, and joking aggression) by tracking escalation, role shifts, and repair refusal across discourse trajectories. We conclude with implications for campus policy, reporting ecosystems, and discourse-informed digital citizenship pedagogy, and we outline an agenda for multilingual, cross-cultural university settings and ethical governance of sensitive online data.

KEYWORDS: applied linguistics, appraisal, CDA, CMDA, corpus-assisted discourse studies, cyberbullying, discourse analysis, impoliteness, multimodality, stance, university students

Research Journal in Advanced Humanities

Volume 6, Issue 4, 2025

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

ARTICLE HISTORY

Submitted: **15 November 2025**

Accepted: **25 November 2025**

Published: 31 **December 2025**

HOW TO CITE

Alowayyid, N. N. S., & Khasawneh, M. A. S. (2025). Discourse analysis of cyberbullying among university students: An applied linguistics approach. *Research Journal in Advanced Humanities*, *6*(4). https://doi.org/10.58256/5epj5m33

Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited

1. Introduction

As social networking sites have spread rapidly across the globe and the majority of users are adults and teenagers, the phenomenon of cyberbullying has begun to be prevalent among such age populations (see. e.g.,Wilson, et, al, 2012; Boyd et, al, 2007; Tokunaga, 2010; Smith, 2008; Patchin, 2006). Statistics have demonstrated that university students spent most of their time on such social networking sites. They demonstrated different types of reactions ranging from love, respect, and hate speech to cyberbullying. (Chandrasena et al., 2022; EDUCAUSE, 2018). Students are unconsciously motivated to indulge in such electronic bullying, whose psychological consequences might be destructive to the majority of the students. (Arif et al., 2024; Alghamdi et al., 2025). PLOS+2PubMed+2They might suffer from mental disorder, depression, stress, or anxiety. (Kaur et al., 2019; Al Serhan & Elareshi, 2019). This phenomenon has been extensively investigated in psychology. However, there are few linguistic studies that attempt to address cyberbullying (Suler, 2004). The present study argues that a linguistic study based on creative methodologies would eventually reveal the invisible reasons for such a problem and why it is common among university students. Therefore, a discourse analysis is adopted as a theoretical framework for decoding the mysterious language used by the aggressors online and how such language is reframed, acquiring different and pejorative and insulting meanings from their formal meanings. The studies conducted in the Arab world addressing the phenomena of university cyberbullying are so inadequate, and such a few number has examined it from a purely psychological perspective that search for the psychological reasons that cause it and bring it to reality.

Cyberbullying among university students is widely framed as a public health risk, a psychological stressor, or a safety problem requiring institutional governance. Those perspectives are considered to significant, as the present study lay emphasis on detecting the associations between university students' engagement in cyberbullying and adverse mental health outcomes (e.g., stress, anxiety, depression) (Bansal et al., 2024). Although these adverse mental health outcomes are investigated from psychological perspectives, they have not been examined yet from a linguistic perspective that focus on **language and semiosis**— tackling online posts, comments, images, emojis, hashtags, quote-posts, and reposting practices. Cyberbullying is therefore not merely something that happens *in* discourse; it is something that happens *as* discourse. A frequently cited definition describes cyberbullying as an aggressive, intentional act carried out via electronic contact, repeatedly and over time, against a victim who cannot easily defend themselves (Smith et al., 2008,). Although providing an explicit definition to cyberbullying is a highly controversial issue, several elements are common among all different tackling the concept ranging from intentionality, repetition, power imbalance or asymmetry, and mediated contexts creating persistence, publicity, and intentional aggression. At the university level, prevalence estimates vary by region, instruments, and sampling, but multiple studies indicate that a substantial minority of students experience victimization and/or perpetration, including overlapping "bully–victim" roles (e.g., nursing students in Jordan; college samples in China). Such findings justify prevention and intervention efforts, but they do not, by themselves, explain how cyberbullying is accomplished interactionally: **how language recruits audiences, amplifies humiliation, legitimizes aggression, and constrains targets' responses.**

Applied linguistics offers robust tools to address those questions. **Computer-Mediated Discourse Analysis** (CMDA) provides a language-centered approach to studying online behavior and how technological features condition discourse (Herring, 2004). **Discourse Analysis** (CDA), especially Fairclough-inspired approaches, supports connecting micro-textual features to discursive practice and broader social structures (as clearly operationalized in Janks, 2005). **Appraisal theory** explains how evaluation, stance, and intensity are encoded lexico-grammatically (Martin & White, 2005). **Facework and (im)politeness** traditions model how interaction threatens, attacks, negotiates, or repairs social identity (face) (Goffman, 1967; Brown & Levinson, 1987) and how impoliteness functions as strategic face-attack often entangled with power (Culpeper, 1996). Despite these resources, research on cyberbullying in higher education tends to emphasize (a) psychological measurement and correlational modeling, (b) survey-based prevalence and predictors, and (c) computational detection of abusive language. Detection work has advanced substantially, including taxonomies of online abusive language and reviews of textual cyber-abuse detection; however, many detection approaches treat "toxicity" as an isolated property of single messages rather than an emergent **interactional trajectory** distributed across participants and time. Conversely, discourse-analytic studies often deliver rich qualitative insight into meaning-making and power but can be criticized for limited scale or reproducibility if methods are not explicitly operationalized and

validated.

This paper responds to that gap by presenting a publishable, explicitly structured, and reproducible discourse-analytic framework—**Discourse Analysis of Cyberbullying Among University Students: An Applied Linguistics Approach**—with two central aims: (1) to advance a novel theoretical synthesis (the **Stance–Face–Power Model**), and (2) to detail a scalable methodology (**CAIR-CDA**) including an operational metric (**DHI**) that can be implemented in full empirical studies under appropriate ethical approvals.

1.1. Contributions

This paper makes four contributions:

1. **Theoretical synthesis:** We develop the **Stance–Face–Power (SFP) Model**, integrating CMDA (Herring, 2004), CDA (Janks, 2005), Appraisal (Martin & White, 2005), and facework/(im)politeness (Goffman, 1967; Brown & Levinson, 1987; Culpeper, 1996) to conceptualize cyberbullying as discursive harm.
2. **Methodological innovation and creativity:** A **Corpus-Assisted and Interactionally-Refined CDA (CAIR-CDA) are introduced**, integrating corpus-assisted discourse studies logic (pattern discovery) with sequential response analysis and multimodal approach.
3. **Operational metric:** the **Dialogic Harm Index (DHI)** is proposed, which is a kind of a transparent, regenerative gauge capturing how harm is built through evaluation intensity, impoliteness strategy severity, audience alignment, and power cues.
4. **Applied implications:** insights and vision are interpreted into implications for the policies followed by the universities, reporting systems, digital citizenship, and moderation—while highlighting ethical governance of sensitive online data.

2. Background: Cyberbullying in University Contexts and the Case for Discourse

Cyberbullying has received sustained global attention, yet university contexts remain conceptually complex for at least three reasons. First, university students are legal adults in many jurisdictions, which can alter reporting practices, institutional responsibilities, and peer norms of accountability. Second, university cyberbullying often unfolds in **semi-public** spaces (campus forums, society pages, confession accounts, group chats with porous boundaries), where visibility and circulation intensify the social consequences of discourse. Third, the university is a site of "high-stakes identity work": competence, belonging, and moral standing are continuously negotiated in academic and social arenas. These features make cyberbullying in higher education particularly amenable—and urgent—for discourse-analytic inquiry.

2.1. Definitions and key criteria

A widely used definition frames cyberbullying as aggressive, intentional behavior enacted via electronic contact repeatedly over time against someone who cannot easily defend themselves (Smith et al., 2008). This definition foregrounds **power asymmetry** and **repetition**. Importantly, repetition in digital environments can include not only repeated acts by an initial aggressor but also the persistent availability and resharing of harmful content by others. In addition, anonymity, publicity, and platform-specific affordances can lower the cost of participation in aggression and broaden the audience for humiliation.

From an applied-linguistics standpoint, definitions matter because they shape what counts as "data" and how harm is operationalized. A purely lexical definition (e.g., presence of slurs) will miss implicit abuse, insinuation, coded language, and crowd dynamics. Detection research explicitly notes that abusive language is diverse and often indirect or contextual. Discourse analysis can therefore complement definitional criteria by showing how cyberbullying is **interactionally achieved**, even when overtly abusive forms are absent.

2.2. Prevalence and consequences in higher education

Recent research continues to document substantial cyberbullying involvement among university or college populations, with patterns differing by context and sampling. For example, in a study of nursing students in Jordan, notable proportions were classified as perpetrators only, victims only, or both, with some forms (e.g., visual/sexual perpetration; verbal/written victimization) reported as prevalent. In a large study of Chinese college students, cyberbullying victimization prevalence was reported at a meaningful rate, with demographic

and behavioural correlates discussed (e.g., online activity). Despite the fact that prevalence is not symmetrical whose results cannot be taken for granted in terms of generalization across different settings, these studies support the claim that cyberbullying is not limited to school-aged populations. However, cyberbullying is also related to university student populations.

Studies conducted on mental health and well-being of the students have emphasized the grave mental problems stemming from cyberbullying's potential impacts. A recent review detects the cyberbullying–mental health studies underscore the fact that there are various studies the addressing the associations cyberbullying experiences to psychological distress and highlights evolving research directions. This matters for discourse analysis because the **discursive mechanisms** of humiliation, exclusion, and identity degradation provide plausible pathways through which online interactions become psychologically consequential.

2.3. Why discourse analysis is necessary (not optional)

Cyberbullying is, fundamentally, a communicative act—and often a communicative campaign. Discourse analysis have major contributions to the current study by addressing the following:

- **Target construction:** how person are situated as direct objects of ridicule, condemnation, or exclusion.
- **Audience recruitment:** how "bystanders" could be indirectly involved in such malpractices through stance and engagement cues.
- **Accountability management:** how the bullying perpetrators attempt to escape his responsibilities by claiming that he is just joking and making fun outsource agency ("people are saying…"), or trying to reframe his bullying through adopting ethical responsibility and claim.
- **Semiotic intensification:** this is related to the electronic images such as emojis, memes, screenshots, and multimodal and their role in remixing intensify evaluation while maintaining acceptable deniability.
- **Escalation and repair:** how sequences of turns transform conflict into sustained harm, and how repair attempts are accepted or refused.

These concerns align directly with CMDA's program to analyze online discourse across linguistic, interactional, and social dimensions in relation to technological conditions. (

3. Theoretical Framework: The Stance–Face–Power (SFP) Model

The **SFP Model** conceptualizes cyberbullying as discursive harm produced through **stance acts** that do facework in ways that enact and reproduce **power**—both local (interactional) and structural (institutional, sociocultural). The model is designed to be (a) theoretically explicit, (b) analytically operationalizable, and (c) compatible with mixed-methods and computational collaboration.

3.1. CMDA: how platform affordances condition discourse

Herring's CMDA explains that computer-mediated discourse can be decoded, interpreted and analysed at different stages levels that includes structure, meaning, interaction management, social phenomena). His model also underscores emphasizing digital advances are not mere "context" but build discourse options and limiations (Herring, 2004). For cyberbullying, CMDA directs attention to:

- **Participation structure:** Who can speak? Who is visible? Who is anonymous? What are the costs of speaking?
- **Temporal structure:** persistence of posts, delayed responses, recirculation of archived screenshots.
- **Semiotic repertoire:** emojis, GIFs, hashtags, mentions, quote-posts, reaction buttons.
- **Genre and norms:** local campus "confession" pages, call-out threads, rumor cycles, "ratio" culture, or private-to-public leakage.

CMDA also highlight that online discourse is not merely written text; however, it is often multimodal, indexical, and shaped by interface design. Such information is significance because because cyberbullying can be reinforced by affordances enabling rapid uptake, easy agreement tokens, and public visibility.

3.2. CDA: linking micro-text to macro-social practice

CDA offers instruments that link micro-linguistic types and patterns to wider value systems, beliefs, faiths social structures and ideologies. Janks (2005) follows Fairclough's three-dimensional logic in the sense that he foregrounds the discourse analysis across three interrelated stages consisting of description, interpretation, and explanation. He also underscores how discourse and power play a major role in our own understanding and interpretation of any given discourse. Investigating the linguistic structure of cyberbullying, one can notice that micro-level features (e.g., modality, pronouns, presupposition, naming practices, metaphor) can regenerate macro-level ideologies and value systems (gender norms, xenophobia, ableism), institutional inequalities, and campus hierarchies.

In SFP, a Fairclough-inspired three-layer analytic orientation is adopted as theortical framework (via Janks, 2005):

1. **Text:** linguistic and multimodal resources used;
2. **Discursive practice:** it addresses different types of texts could be produced and reproduced, disseminated consumed (commenting, screenshotting, resharing);
3. **Social practice:** how wider power relations and ideologies are reinforced.

3.3. Appraisal: evaluation, alignment, and intensity

Appraisal theory (Martin & White, 2005) provides a well-organized and a systematic framework for understanding, interpreting, assessing language and linguistic structures, ranging from **Attitudinal modes such affective, judgmental, aprpeactive modes** (affect, Engagement modes such (monoglossia/heteroglossia; dialogic positioning), to **Graduational modes such as** force/focus; scaling intensity). Cyberbullying frequently relies on:

- **Judgment** (moral/social condemnation): attacks on ethics, normality, competence.
- **Affect:** invoking shame, fear, contempt, disgust.
- **Graduation:** intensifiers, repetition, typographic emphasis, extreme quantifiers.
- **Engagement:** aligning with presumed audiences ("everyone knows…"), rumor framing ("people are saying…"), disclaimers ("no offense but…").

Appraisal is especially valuable for distinguishing generic negativity from **socially consequential evaluation** that licenses exclusion or humiliation.

3.4. Facework and (im)politeness: harm as strategic face-attack

Goffman's notion of face conceptualizes the positive social value people claim through interaction, with facework as the practices used to maintain or repair that value (Goffman, 1967). Brown and Levinson's politeness theory models how speakers mitigate face-threatening acts in interaction (Brown & Levinson, 1987). In contrast, Culpeper's impoliteness framework systematically theorizes communicative strategies oriented toward attacking face and causing disharmony, including mock politeness and contextual dynamics (Culpeper, 1996).

Cyberbullying is often characterized by sustained face-attacks that can include:

- **Bald-on-record attacks:** direct insults, commands, threats.
- **Positive-face attacks:** humiliating the desire for approval (mocking competence, appearance, morality).
- **Negative-face attacks:** coercion, demands, control of participation ("stop talking"), intimidation.
- **Mock politeness/sarcasm:** "with all due respect…," "sweetie," "no offense but…" weaponized.
- **Withholding politeness/uptake:** ignoring repair, dismissing clarification, denying the right to respond.

The significance of linking facework with evaluation, appraisal and stance, SFP is capable of not only capturing what is said but what is done to social identity, involvement and belonging.

3.5. Stance: evaluation, positioning, and alignment

Stance is defined as a kind of social reaction that covers different aspects like evaluation, positioning, and alignment. Du Bois's "stance triangle" includes three parts, simultaneously a stance object, positioning a subject (self/other), and aligning with other subjects—focusing on the dialogic characteristics of stance in interaction (Du Bois, 2007).) In cyberbullying, stance-taking is central to:

- Attracting bystanders into agreement and amplification;
- Making binary opposition between "us vs. them
- Reducing the human being with sensations and feelings to stance objects (memes, rumors, stereotypes) rather than persons.

3.6. The SFP model: an integrated proposition

SFP proposition: Cyberbullying among university students is discursive harm produced through **stance acts** that (a) attack face and (b) mobilize power under platform affordances, scaling harm across time and audiences.

- **Stance** explains how evaluation and alignment are organized (Du Bois, 2007).
- **Face** gives explanations whey specific movements could interpreted as socially disheartening (Goffman, 1967; Brown & Levinson, 1987; Culpeper, 1996).
- **Power** explains who can attack with a feeling that **he will not be held accountable to punishment or consequences**, especially when someone does something wrong and is **not held accountable** to something wrong he does, and how institutional/platform structures magnify consequences (Janks, 2005; Herring, 2004).

4. Research Questions

Guided by SFP, this paper advances the following research questions (RQs), structured for full empirical implementation:

RQ1 (Discursive resources): What recurrent linguistic and multimodal resources (lexical, grammatical, pragmatic, visual) realize cyberbullying in university-affiliated online interaction?

RQ2 (Trajectory): How does discursive harm escalate across interactional sequences (initiation → uptake → piling-on → institutionalization), and what discourse moves predict de-escalation or repair?

RQ3 (Alignment and roles): How are participant roles (perpetrator, target, defender, assistant, outsider/reporter) enacted through stance and engagement resources, and how do audiences become co-producers of harm? (cf. role-sensitive cyberbullying analyses).

RQ4 (Power and ideology): How do cyberbullying episodes reproduce or contest social hierarchies (gendered, racialized, classed, ableist) and institutional power relations on campus?

RQ5 (Applied implications): How can discourse-analytic findings inform interventions in policy, reporting, pedagogy, and moderation?

5. Methodology: Corpus-Assisted, Interactionally-Refined CDA (CAIR-CDA)

This section details a method designed for Q1/WoS-level expectations: explicit operationalization, reproducibility, and triangulation. CAIR-CDA can be implemented as a full empirical study when ethical approvals and data governance are secured. Here, we present the method and demonstrate analytic logic using sanitized composite vignettes (Section 6) to avoid reproducing harm or exposing identities.

5.1. Design overview

CAIR-CDA is a three-stage pipeline:

1. **Corpus-assisted mapping (pattern discovery):** Identify high-frequency patterns, keywords, collocations, and evaluative bundles associated with bullying episodes compared with control discourse (non-bullying campus talk). This step supports scale and reduces analyst cherry-picking.
2. **Sequential interaction analysis (CMDA layer):** interpret episodes step by step: how posts evokes responses, how responses align or resist" **how those features help a message (including harmful content) spread viral , reinforce , and attract more harmful participation.** (Herring, 2004). (
3. **Critical explanation (CDA layer):** Explain how repetitive micro-patterns are associated to a wider social practices and institutional conditions (Janks, 2005).

This design is intentionally "two-way": corpus patterns guide qualitative deep dives, and qualitative insights refine corpus queries and annotation categories.

5.2. Data sources and corpus construction (for full empirical studies)

A publishable empirical CAIR-CDA project can draw on three ethically defensible corpora types:

Corpus A: University-affiliated public online discourse

Public posts and comment threads from university communities (e.g., public campus forums, student society pages, campus confession pages where permitted by platform terms). Inclusion criteria might require (i) clear university relevance (course, dorm, campus event), (ii) interpersonal targeting (identifiable person or socially indexable target), and (iii) interaction (responses, quote-posts, reposting).

Corpus B: Institutional discourse

University releases harassment codes, harassment policies that provide guidance, and awareness campaigns. These texts provide official definitions of harm and responsibility , offering discursive comparison with student discourse.

Corpus C: Moderation/repair discourse:

Online threads that follows post plus responses and comments as the defender tries to stop the harm and the problems or how cyberbullying can be resisted or supported through discourse.

Ethical governance:

Cyberbullying research overlaps with psychological and mental health risk and potential re-identification. Researchers should lay emphasis on prioritize anonymization, careful paraphrase when quoting, secure storage, and strict minimization of sensitive content.

5.3. Unit of analysis: The Cyberbullying Episode (CBE)

A **Cyberbullying Episode (CBE)** is a bounded interactional sequence (single post or multi-turn thread) that includes:

- A targeting move (explicit or inferable);
- At least one harmful stance act (ridicule, humiliation, threat, dehumanization, exclusion, exposure);
- Evidence of repetition, escalation, or audience involvement—aligned with definitional criteria emphasizing repeated aggression and constrained defense.

In CAIR-CDA, a CBE is not defined by a single "toxic" utterance but by a **trajectory**: how talk unfolds, how others align, and how repair is treated.

5.4. Annotation scheme: SFP coding layers

Each CBE can be annotated across four layers, with a codebook and exemplars.

Layer 1: Stance–Appraisal (Martin & White, 2005)

- Attitude: Affect / Judgment / Appreciation
- Engagement: attribution, disclaimers, dialogic contraction/expansion
- Graduation: force/focus, intensification, repetition.

Layer 2: Face–(Im)politeness (Goffman; Brown & Levinson; Culpeper)

- Strategy type: bald attack, sarcasm/mock politeness, dismissals, threats, tabooing, exclusion cues
- Face target: positive vs negative face
- Repair: apologies, clarifications, meta-pragmatic comments; acceptance vs refusal

Layer 3: Power cues

- Interactional power: commands, entitlement claims, gatekeeping ("you don't belong"), exposure/ screenshot threats
- Structural power: appeals to institutional authority, group dominance, ideologies (gendered/ableist/ xenophobic positioning)

Layer 4: Platform and multimodality

- Mentions/tags, quote-posts, screenshotting, emoji sequences, memes, image macros
- Where visual data are analyzed, multimodal discourse traditions (e.g., visual grammar) can inform annotation of gaze, framing, salience, and composition (Kress & van Leeuwen, 2006/2020).

5.5. A distinctive contribution: The Dialogic Harm Index (DHI)

Computational metrics often reduce harm to toxicity scores; behavioral measures often miss interactional scaling. DHI is designed to be discourse-sensitive and transparent.

For each turn (t) within a CBE:

- **E(t): Evaluative load** (Appraisal polarity/type; Judgment typically weighted higher than Appreciation; Affect weighted for shame/fear cues)
- **G(t): Graduation intensity** (caps, repetition, intensifiers, extreme quantifiers)
- **I(t): Impoliteness severity** (threats/exposure > direct insult > mock politeness > dismissals)
- **A(t): Alignment amplification** (piling-on: agreement tokens, echoing, "ratio" logic, meme repetition)

[DHI(CBE) = \sum_{t=1}^{T} \big(w_E E(t) + w_G G(t) + w_I I(t) \big)\times (1 + w_A A(t))]

Weights (w) are set a priori (registered) and/or calibrated via validation (e.g., correlation with moderation outcomes, or—where ethically possible—target-reported distress). The key conceptual move is that **alignment amplifies harm**: small "lol/true/this" responses may contain little toxicity lexically but can increase social proof and persistence.

5.6. Reliability, validity, and transparency

To meet high-impact journal expectations, CAIR-CDA should report:

- **Inter-annotator agreement:** Cohen's κ / Krippendorff's α for key categories (Appraisal, impoliteness, role codes)
- **Audit trail:** codebook versions, annotated exemplars, analytic memos
- **Triangulation:** corpus patterns → sequential analysis → CDA explanation (multi-level convergence)
- **Negative case analysis:** include borderline interactions (banter, conflict, roasting) to sharpen distinctions

This combination aims to preserve interpretive depth while increasing reproducibility.

6. Demonstration: What CAIR-CDA Reveals (Sanitized Composite Vignettes)

Transparency statement: The following vignettes are *constructed composites* designed to explain the analytical framework without regenerating harmful content or revealing identities. They are not presented as experimental findings from a specific dataset.

6.1. Pattern 1: Moralizing call-outs as judgment + audience recruitment

A call-out is regarded as cyberbullying when the aggressor frames the target as a bad person who never follows moral code and then forces other people to accept his viewpoint and join him in cyberbullying. It is not just criticism; however, it

Trajectory: accusation → moral judgment → coercive alignment → identity damage

Vignette (sanitized):

> Post: "So we're just going to ignore what X did in the lab? Unbelievable."
> Comment: "Exactly. People like that don't belong here."
> Comment: "If you defend them, you're part of the problem."
> Comment: "Screenshots don't lie."

SFP analysis

- **Appraisal:** Dominant *Judgment* constructs the target as morally defective, not merely mistaken; Graduation intensifies certainty ("unbelievable," categorical belonging claims). (Martin & White, 2005)

- **Engagement:** Presupposition positions agreement as default and dissent as deviance ("we're just going to ignore…").

6.2. Pattern 2: Mock politeness and plausible deniability
Vignette:

Comment: "With all due respect, you should probably stop talking in class."
Reply: "That's rude."
Comment: "Not rude—just honest. Don't be so sensitive."

SFP analysis

- **Impoliteness:** Mock politeness functions as a face-attack with a civility "shield" (Culpeper, 1996).
- **Engagement:** "Just honest" frames the speaker as truth-teller; it delegitimizes the target's meta-pragmatic protest.
- **Facework:** Repair attempt ("rude") is refused, escalating harm by denying the target's right to define the act.
- **Power:** The attacker claims entitlement to regulate participation ("stop talking"), a negative-face attack (Brown & Levinson, 1987).

Analytic payoff: Discourse analysis reveals how "polite" framings can intensify harm by enabling deniability and shifting blame to the target ("too sensitive").

6.3. Pattern 3: Negation and gatekeeping ("You're not…")
Vignette:

"You're not a real student if you need accommodations."
"You're not even from here."

SFP analysis

- **Appraisal:** Judgment targets normality/legitimacy; negation performs identity denial.
- **Power:** Gatekeeping attacks institutional belonging and can reproduce ideologies (ableism/xenophobia) in campus life.
- **Detection relevance:** Abusive-language research highlights the need to capture implicit abuse beyond explicit slurs.

Analytic payoff: The harm is partly in what is *presupposed* and what identities are made "unsayable" or illegitimate—not merely in swearwords.

6.4. Pattern 4: Institutional discourse vs student discourse (interdiscursivity)
Vignette:

Student: "Report me then "
Policy-style talk: "Harassment includes repeated conduct that creates a hostile environment…"

SFP analysis

- **Engagement:** Humor anticipates and neutralizes institutional threat; it frames reporting as unserious.
- **Power:** Institutional authority is invoked yet delegitimized; this reveals a gap between policy language and student uptake.
- **CDA:** Interdiscursivity matters because institutional discourse may fail when it does not recognize how students strategically reframe accountability talk (Janks, 2005).

7. Discussion

7.1. Cyberbullying as discursive harm: what SFP adds

SFP reframes cyberbullying as a process in which harm emerges through three interlocking mechanisms:

1. **Evaluative stance (Appraisal):** negativity is not merely sentiment; it is moral/social positioning that licenses exclusion and humiliation (Martin & White, 2005).
2. **Face-attack (im)politeness:** cyberbullying is interactional aggression that damages identity claims and often blocks repair (Goffman, 1967; Culpeper, 1996).
3. **Power mobilization:** harm scales with audience alignment, platform affordances, and institutional contexts (Herring, 2004; Janks, 2005).

This triad clarifies why cyberbullying can be severe even when individual messages appear mild: harm emerges from **trajectory + audience + persistence**.

7.2. Distinguishing cyberbullying from conflict, critique, and joking aggression

University life includes conflict, critique, sarcasm, and "banter." A discourse-informed approach distinguishes cyberbullying by tracking:

- **Targeting under asymmetry:** Is the target positioned as unable to respond safely?
- **Persistence/recirculation:** Are screenshots/reposts used to sustain harm?
- **Audience recruitment:** Are others compelled to align ("if you defend them…") and does the crowd amplify?
- **Repair refusal:** Are apologies/clarifications rejected, and is meta-pragmatic negotiation blocked?
- **Role fluidity:** Do bystanders shift into assistants/defenders/reporters?

Role-sensitive work shows that cyberbullying dynamics involve multiple roles beyond bully–victim binaries, supporting the importance of analyzing participation structures and alignment.

7.3. Why corpus assistance improves rigor (without losing interpretive depth)

Corpus-assisted mapping can reduce selective attention by highlighting "non-obvious" patterns (e.g., recurring moralizing bundles, deniability markers, or gatekeeping phrases). This does not replace qualitative analysis; it guides it. Importantly, CAIR-CDA treats corpus outputs as *hypothesis-generating*—to be confirmed, refined, or challenged through sequential analysis and CDA explanation.

7.4. Bridging discourse analysis and computational detection

Abusive-language detection research has made major progress, including taxonomies and reviews emphasizing variability (explicit/implicit abuse, context dependence). Yet many models struggle with sarcasm, coded language, or harm emerging through piling-on. CAIR-CDA provides a bridge: detection can flag candidate episodes, while discourse analysis determines trajectory, role dynamics, and explainable harm mechanisms. DHI can serve as an interpretable supervisory signal or evaluation criterion for systems that aim to capture interactional harm, not only toxicity.

8. Applied Implications for Universities

8.1. Policy: from legalistic definitions to interactional indicators

University policies are often legally cautious and procedurally oriented. Discourse analysis suggests adding interactional indicators that students recognize as "what actually happens," such as:

- sustained screenshot circulation and recontextualization;
- coercive alignment threats ("if you defend them…");
- gatekeeping language that denies belonging;
- repeated mock politeness used to legitimize face-attack;
- crowd amplification via reactions and echoing.

Such indicators do not replace legal definitions; they operationalize early-warning signs and training materials.

8.2. Reporting ecosystems: supporting targets and reducing amplification

Because repetition and persistence are core features of cyberbullying, reporting systems should make it easy to capture episode-level evidence (thread context, timestamps, platform features). Support services can also provide discourse-informed guidance on response strategies that reduce amplification (e.g., avoiding engagement that feeds the thread; documenting; seeking defender support). This matters given documented mental-health stakes in cyberbullying research.

8.3. Pedagogy: digital citizenship as discourse competence

Many interventions are framed as awareness. A discourse approach reframes prevention as **communicative competence**:

- teaching how stance and engagement recruit audiences;
- recognizing mock politeness and deniability;
- practicing defender discourse (supportive alignment without escalation);
- developing norms for disagreement that preserve face and belonging.

Applied linguistics is well positioned to design and evaluate such curricula because it can connect micro-linguistic choices to social effects.

8.4. Moderation and community governance

Moderation policies often focus on prohibited words. CAIR-CDA suggests episode-based moderation cues:

- rapid alignment cascades (many "this/facts/lol");
- repeated naming/tagging;
- explicit calls to expose or circulate content;
- refusal of repair and escalation markers.

DHI can offer a transparent way to explain why a thread is harmful even when individual posts are not overtly abusive.

9. Conclusion and Research Agenda

This paper argued that cyberbullying among university students should be analyzed as **discursive harm**—a patterned interactional accomplishment shaped by stance-taking, face-attack, and power. The proposed **Stance–Face–Power (SFP) Model** integrates CMDA (Herring, 2004), CDA (Janks, 2005), Appraisal (Martin & White, 2005), and facework/(im)politeness (Goffman, 1967; Brown & Levinson, 1987; Culpeper, 1996), yielding a coherent applied-linguistics approach with clear analytical and applied payoffs. Methodologically, **CAIR-CDA** offers a scalable pipeline combining corpus mapping with sequential analysis and critical explanation, while the

Dialogic Harm Index (DHI) operationalizes harm as emergent interaction rather than isolated toxicity.

Future research should extend CAIR-CDA in five directions:

1. **Multilingual university ecologies:** code-switching and language ideologies in cyberbullying.
2. **Cross-cultural comparisons:** how face, shame, honor, and belonging are discursively constructed in different higher-education contexts.
3. **Longitudinal trajectories:** how online episodes reshape on-campus participation and identity over time.
4. **Intervention evaluation:** whether discourse-based defender training reduces piling-on and improves repair outcomes.
5. **Human–AI moderation:** integrating explainable, episode-based metrics (like DHI) with detection pipelines that recognize implicit and interactional harm.

By making discourse central, applied linguistics contributes not only to explanation but also to actionable tools for prevention, community repair, and institutional governance.

Data Availability Statement
No empirical dataset is provided in this framework-and-methods paper. Researchers implementing CAIR-CDA should publish codebooks, annotation guidelines, and derived frequency tables where ethically permissible and provide controlled access procedures for sensitive data.

Funding:
The author extends their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the Project number (PSAU/2025/02/36984).

Conflict of Interest Statement
The author(s) declare no conflict of interest.

References

Al Serhan, F., & Elareshi, M. (2019). University students' awareness of social media use and hate speech in Jordan. *International Journal of Cyber Criminology, 13*(2), 548–563. bibliomed.org+1

Al-shatnawi, F. E., Ababneh, A. M. T., & [additional authors]. (2024). Prevalence of cyberbullying perpetration and cyberbullying victimization among nursing students in Jordan. *SAGE Open Nursing, 10*, 23779608241256509. https://doi.org/10.1177/23779608241256509 PMC

Alghamdi, W., & [additional authors]. (2025). Relationship between cyberbullying, anxiety, and depression among university students in Jeddah, Saudi Arabia: A cross-sectional study. *International Journal of Psychiatry in Medicine, 60*(2), 170–186. PubMed

Arif, A., Qadir, M. A., Martins, R. S., & Khuwaja, H. M. A. (2024). The impact of cyberbullying on mental health outcomes amongst university students: A systematic review. *PLOS Mental Health, 1*(5), e0000166. PLOS

Bansal, S., Garg, N., Singh, J., & van der Walt, F. (2024). Cyberbullying and mental health: Past, present and future. *Frontiers in Psychology, 14*, 1279234. https://doi.org/10.3389/fpsyg.2023.1279234 Frontiers+1

boyd, d. m., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x OUP Academic+1

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.

Chandrasena, P. P. C. M., & Ilankoon, I. M. P. S. (2022). The impact of social media on academic performance and interpersonal relations among health sciences undergraduates. *Journal of Education and Health Promotion, 11*, 117. https://doi.org/10.4103/jehp.jehp_603_21 PMC

Culpeper, J. (1996). Towards an anatomy of impoliteness. *Journal of Pragmatics, 25*(3), 349–367. https://doi.org/10.1016/0378-2166(95)00014-3 ScienceDirect+1

Díaz-García, J. A., & Carvalho, J. P. (2025). A literature review of textual cyber abuse detection using cutting-edge natural language processing techniques: Language models and large language models. *WIREs Data Mining and Knowledge Discovery*, e70029. https://doi.org/10.1002/widm.70029 Wires+1

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139–182). John Benjamins. https://doi.org/10.1075/pbns.164.07du ResearchGate+1

EDUCAUSE Center for Analysis and Research. (2018). *ECAR study of undergraduate students and information technology: 2018* (section: "A day in the online life of a student"). EDUCAUSE. EDUCAUSE+1

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. Pantheon Books. eClass UOA+1

Guo, Y., Cai, S., & Liang, F. (2024). The prevalence of cyberbullying and the association between cyberbullying, emotional exhaustion, and perception of collective efficacy among Chinese college students. *Journal of Affective Disorders, 362*, 145–151. https://doi.org/10.1016/j.jad.2024.06.113 PubMed

Herring, S. C. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In S. A. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for virtual communities in the service of learning* (pp. 338–376). Cambridge University Press. https://doi.org/10.1017/CBO9780511805080.016 Cambridge University Press & Assessment+1

Janks, H. (2005). Language and the design of texts. *English Teaching: Practice and Critique, 4*(3), 97–110. hilaryjanks+1

Kaur, W., Balakrishnan, V., Rana, O., & Sinniah, A. (2019). Liking, sharing, commenting and reacting on Facebook: User behaviors' impact on sentiment intensity. *Telematics and Informatics, 39*, 25–36. https://doi.org/10.1016/j.tele.2018.12.005 ORCA+1

Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (3rd ed.). Routledge. https://doi.org/10.4324/9781003099857 Find Researcher+1

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan. Prrwhite+1

Moghaddam, S. H., Lyons, K., Regehr, C., Goel, V., & Regehr, K. (2025). *Towards a comprehensive taxonomy of online abusive language informed by machine learning* (arXiv:2504.17653). arXiv. arXiv

Patchin, J. W., & Hinduja, S. (2006). Bullies move beyond the schoolyard: A preliminary look at cyberbullying. *Youth Violence and Juvenile Justice, 4*(2), 148–169. https://doi.org/10.1177/1541204006286288 SAGE

Journals+1
Sheikh, M. M. R., Hossan, R., & Menih, H. (2023). Cyberbullying victimization and perpetration among university students in Bangladesh: Prevalence, impact and help-seeking practices. *Journal of School Violence, 22*(3), 1–17. https://doi.org/10.1080/15388220.2023.2168681 PMC+1
Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry, 49*(4), 376–385. https://doi.org/10.1111/j.1469-7610.2007.01846.x ACAMH Online Library+1
Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7*(3), 321–326. https://doi.org/10.1089/1094931041291295 John Suler+1
Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior, 26*(3), 277–287. https://doi.org/10.1016/j.chb.2009.11.014 ScienceDirect+1
Wilson, R. E., Gosling, S. D., & Graham, L. T. (2012). A review of Facebook research in the social sciences. *Perspectives on Psychological Science, 7*(3), 203–220. https://doi.org/10.1177/1745691612442904 SAGE Journals+1
Zhong, J., Mo, Y., Zhang, J., Liu, P., Luo, X., Liu, L., Ding, R., Huang, J., & Zheng, Y. (2025). Beyond anger: Uncovering complex emotional patterns between cyberbullying roles through affective computing and epistemic network analysis. *Humanities and Social Sciences Communications, 12*, Article 1281. https://doi.org/10.1057/s41599-025-05689-9